Prediction of Humidity using Multi Linear Perceptron and Support Vector Machine North Twenty-four Parganas, West Bengal

Rupam Sardar[1], Sudip Ghosh[2],Bimal Datta[3]

[1]Budge Budge Institute of Technology, Kolkata – 700137,India

[2]Indian Institute of Engineering Science and Technology, Botanical Garden Area, Howrah, West Bengal 711103

[3]Budge Budge Institute of Technology, Kolkata – 700137, India

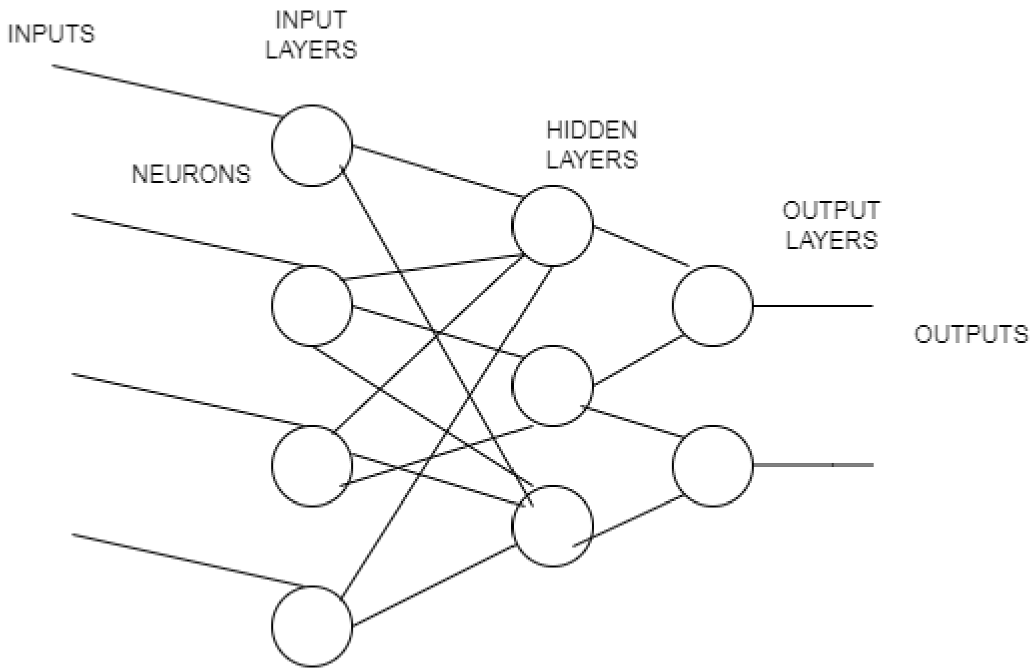
Abstract:

The ability to predict the weather is necessary for lifesaving. Weather prediction has important applications in transportation, agriculture, floods, and other natural phenomena. For weather forecasting to be accurate, air pressure prediction is essential. Among the methods available for Air Pressure prediction is multiple linear regression. To ascertain the relationship between one independent variable and one dependent variable The linear regression model is a very useful tool in statistics. An expansion of the linear regression model is multiple linear regression. To add variables to our experiment, we used MATLAB. Predicting air pressure is crucial to weather forecasting. We have selected Kolkata, or more precisely the Dum Dum Metropolitan area in North 24 Parganas, for the purpose of forecasting rainfall.

Keyword: Humidity prediction, Multi Layer Perceptron Model, Support Vector Machine,Root Mean Square

Introduction:

For plants, animals, and people to exist in a certain location, the air pressure of the earth is necessary.The greatest and minimum air pressures in a climate determine how hot or cold it is. We use a linear regression model to determine the maximum and lowest Air Pressure.(7) Using linear regression is simple when measuring a quantifiable outcome. Assumptions regarding values between the observed values and explanatory factors can be made using this explanatory variable. The least squares approach can be used to fit a linear regression model and identify the best-fitting data. Y is the dependent variable and x is the independent variable in a simple regression model. MLP is another word for multi-layer perception.



Multilayer mathematical model of Perceptrons

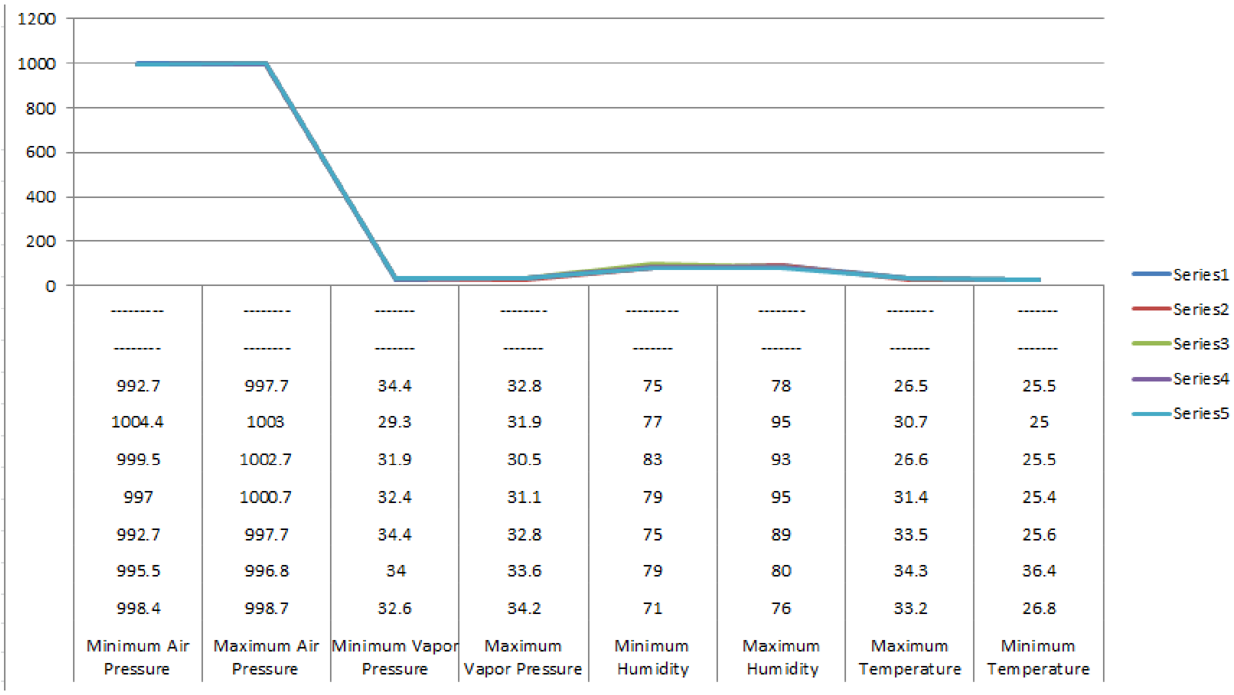
w1x1 + w2x2 + ⋯ + wnxn − θ =

The main goal of the study is to identify the Dum Dum region's highest and lowest air pressures.[3] We may use the least squares method to draw a line with data points that represents the air pressure in the Dum Dum municipality.[1]Simple linear regression models fit in a straight line, but multilayer layer perceptrons fit in a plane. This is how the two types of linear regression models are different. Each independent variable's coefficient is incorporated into the Multi Layer Perceptron Model, and the regression coefficient is obtained in step two using the least squares method.
The linear regression model offers one illustration of how various input models could result in a single output model. Machine learning and statistics are two domains where linear regression models find application.The formula for a basic regression model with the dependent variables x and y is Y=B0+B1X. We can compute the air pressure in the Dum Dum meteorological area by using Matlab to create a linear regression model.
Conversely, when Multi-Layer Percepotron is being used, the equation will be wnxn − θ = w1x1 + w2x2 + ⋯ + A multiple-layer artificial neural network technique is represented by the multi-layer perceptron (MLP). Differentially linear situations can be handled by a single perceptron; however, non-linear cases are not well suited for this method.
In this case, Y represents the maximum and lowest air pressures. The input matrix X represents the highest and lowest values of humidity, vapor pressure, air pressure, and air pressure. Next, we use B to compute the coefficient. Finding the anticipated maximum and lowest air pressures for the following day is the next stage.We learn how the real and expected air pressures vary day by day. Moreover, there are obvious mistakes occurring. A=y*X, B=K*y*Y, y=X', and K=A^-1
A mathematical relationship between several random variables can be discovered using the Multi-Layer Perceptron (MLP).
w1 + w2 + ⋯ + wnxn − θ = 0 for w1, w2, w3,,...The coefficients are wn, and the input variables are x1, x2,..., xn. The maximum air pressure, minimum air pressure, maximum humidity, minimum humidity, maximum vapor

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

pressure, minimum vapor pressure, maximum temperature, and minimum temperature are the eight factors that we selected.

Minimum Air Pressure	Maximum Air Pressure	Minimum Vapor Pressure	Maximum Vapor Pressure	Minimum Humidity	Maximum Humidity	Maximum Temperature	Minimum Temperature
998.4	998.7	32.6	34.2	71	76	33.2	26.8
995.5	996.8	34.0	33.6	79	80	34.3	36.4
992.7	997.7	34.4	32.8	75	89	33.5	25.6
997.0	1000.7	32.4	31.1	079	095	31.4	25.4
999.5	1002.7	31.9	30.5	83	93	26.6	25.5
1004.4	1003.0	29.3	31.9	077	095	30.7	25.0
992.7	997.7	34.4	32.8	75	78	26.5	25.5
--------	--------	-------	-------	-------	-------	-------	-------
---------	--------	-------	--------	---------	--------	--------	-------
998.5	1003.0	29.3	31.9	077	095	31.4	25.4
997.5	999.0	33.6	25.5	82	90	26.4	24.2
996.2	996.4	32.6	33.0	96	87	31.0	25.6
993.9	996.7	33.4	33.4	86	86	31	25
996.7	1001.2	32.8	33.8	78	80	32.6	25.2



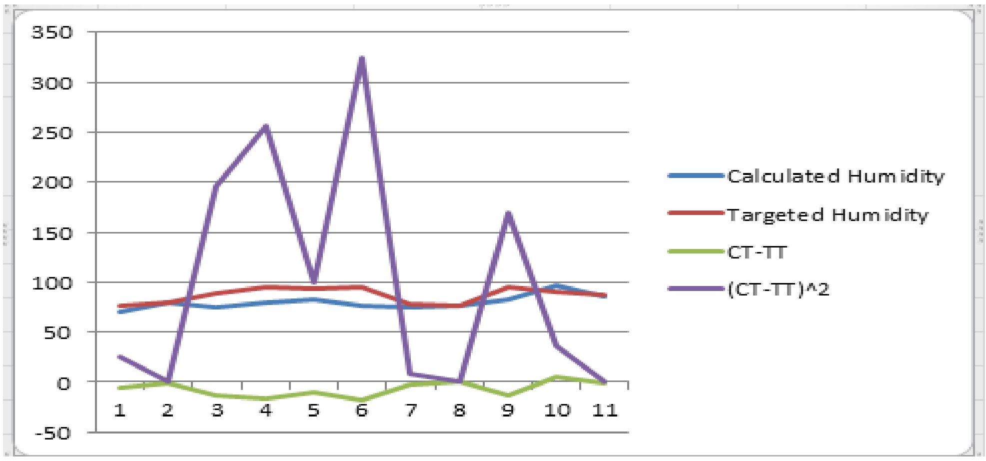
Calculated Minimum Air Pressure:

Calculated Humidity	Targeted Humidity	CT-TT	(CT-TT)^2
71	76	-5	25
79	80	-1	1
75	89	-14	196
079	095	-16	256
83	93	-10	100
077	095	-18	324
75	78	-3	9
077	76	1	1
82	095	-13	169
96	90	6	36
86	87	-1	1
71	76	-5	25
TOTAL			1143

Now Root Mean Square Formula

√ (CT-TT)^2/(N)

=9.759



The mathematical Formula $w_1x_1 + w_2x_2 + \cdots + w_nx_n - \theta = 0$

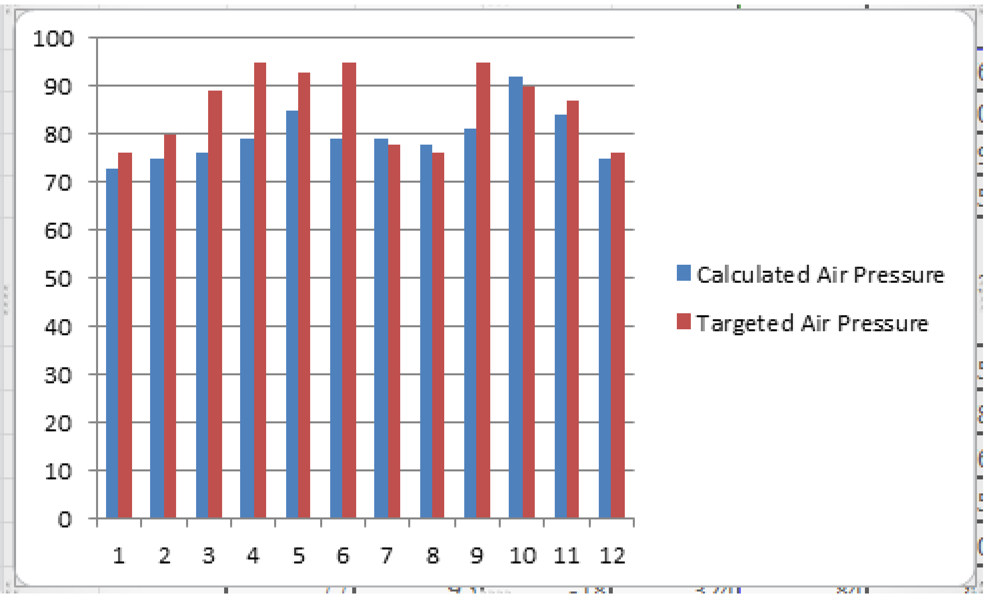
Calculated Humidity	Targeted Humidity
73	76
75	80
76	89
079	095
85	93
079	095
79	78
078	76
81	095
92	90
84	87
75	76

Total=994	

Now Root Mean Square Formula

√ (CT-TT)^2/(N)

=9.101



So for a two dimensional field we can think the targeted and calculated Air Pressure is

Figure 2: Calculated and Targeted Minimum Air Pressure

Calculated Maximum Humidity:

Calculated Humidity	Targeted Humidity	CT-TT	(CT-TT)^2
71	76	-1.9	3.61
79	80	0.9	0.81
75	89	3	9
079	095	2	4
83	93	0.3	0.09
077	095	-5.3	28.09
75	78	5.3	28.09
077	76	-4	16
82	095	-2.6	6.76

96	90	0.3	0.09
86	87	4.5	20.25
78	86	1.5	2.25
TOTAL			119.04

Now Root Mean Square Formula =

$\sqrt{(CT-TT)^2/(N)}$

=3.149

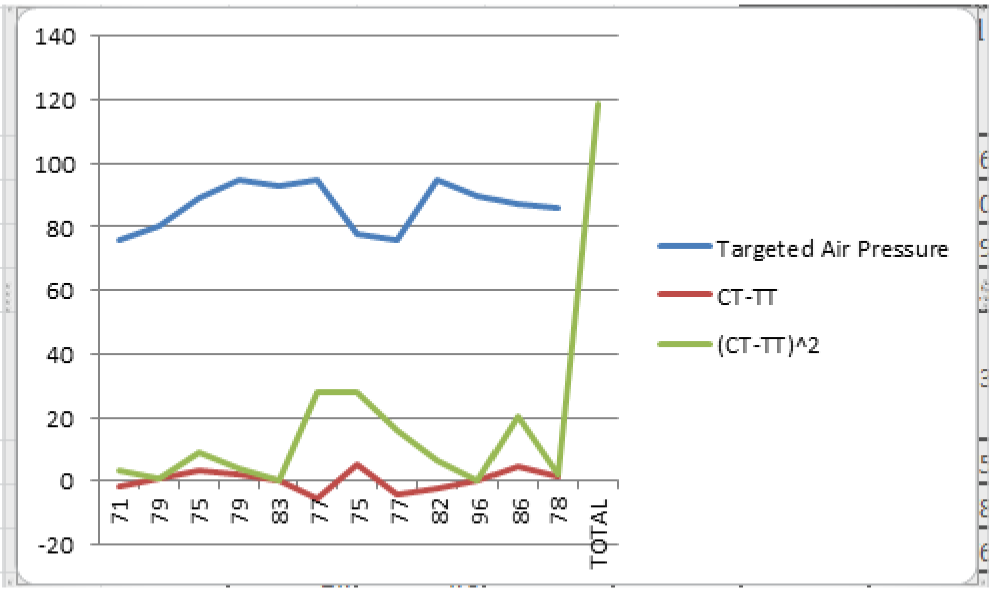


Fig3: Calculated Highest Humidity

The Support Vector Machine (SVM) learning algorithm is presented in this series of notes. SVMs are some of the greatest "off-the-shelf" supervised learning algorithms—many people even argue that they are the best. In order to explain the SVM tale, we must first discuss margins and the notion of dividing data with a big

"gap." We will next stray into the topic of Lagrange duality as we discuss the optimal margin classifier. The SMO technique, which provides a fast SVM implementation, will wrap up the discussion. We will also see kernels, which provide a means of applying SVMs in very high dimensional (such as infinitedimensional) feature spaces.
To begin our story on SVMs, let's discuss margins. The intuitions regarding margins and the "confidence" of our forecasts are provided in this section; Section 3 will formalize these concepts.
Take logistic regression as an example, where hθ(x) = g(θ T x) models the probability p(y = 1|x; θ). Then, if and only if hδ(x) ≥ 0.5, or equivalently, if and only if δT x ≥ 0, we would predict "1" on an input x. Think about an example of positive training (y = 1). As θT x increases, so does hθ(x) = p(y = 1|x; w, b), and hence, our level of "confidence" that the label is 1 increases as well. Therefore, we might consider our forecast to be highly confident informally..

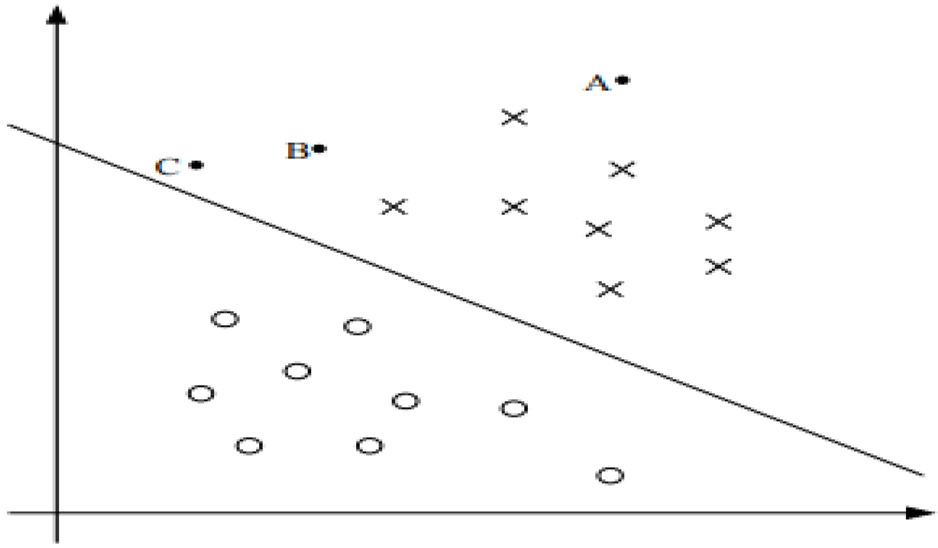


Figure 1:Support Vector Machine

The creation of strategies and tactics that allow computers to learn is the focus of machine learning, which is regarded as a branch of artificial intelligence. Put simply, the creation of algorithms that allow a computer to learn and carry out operations.

1992 saw the introduction of Support Vector Machine (SVM) by Boser, Guyon, and Vapnik at COLT-92. A group of comparable supervised learning techniques used for regression and classification are called support vector machines (SVMs) [1]. They are a member of the generalized linear classifier family. Stated differently, Support Vector Machine (SVM) is a machine learning theory-based classification and regression prediction tool that maximizes predictive accuracy while automatically preventing over-fitting to the data. Systems that use the hypothesis space of a linear function in a high-dimensional feature space and are trained using an optimization theory learning algorithm that applies a statistical learning theory learning bias are known as support vector machines.

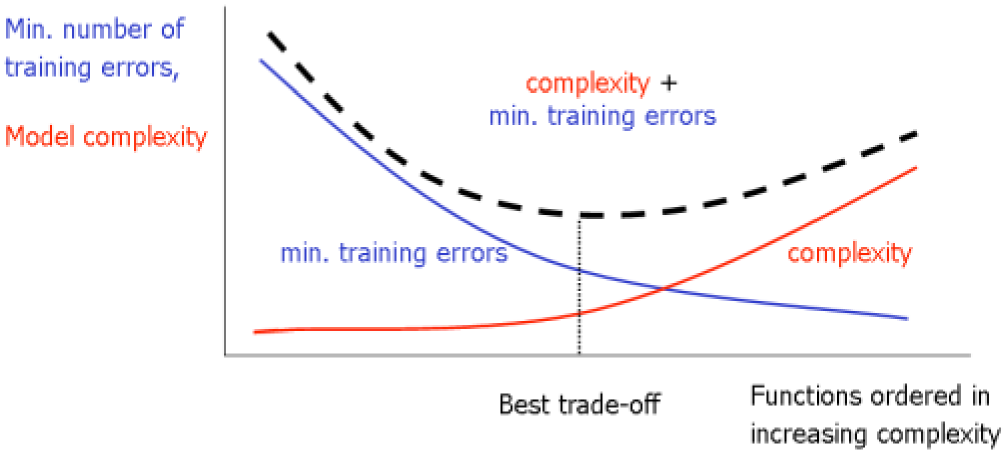
Methodology:

In order to facilitate our explanation of SVMs, it will be necessary to first establish a new notation for classification. In order to solve a binary classification problem with labels y and features x, we will be examining a linear classifier.

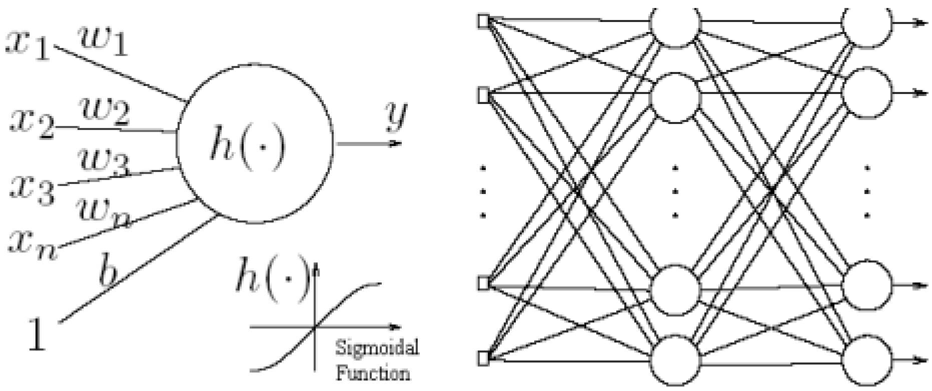
From here on, we will indicate the class labels with $y \in \{-1, 1\}$ rather than $\{0, 1\}$.

Additionally, we will use parameters w, b and express our classifier as $h_{w,b}(x) = g(w^T x + b)$ instead of parameterizing it with the vector θ.

Here, if $z \geq 0$, then $g(z) = 1$, and if not, then $g(z) = -1$. We are able to explicitly consider the intercept term b independently of the other parameters thanks to this "w, b" notation. (We also abandon our earlier practice of allowing $x_0 = 1$.)

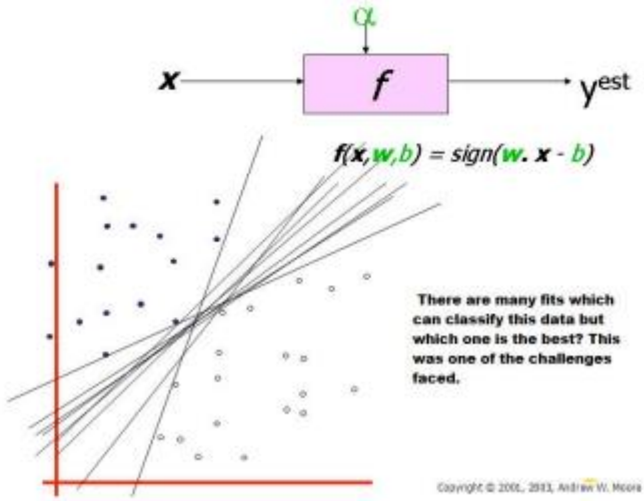


First, employing neural networks for both supervised and unsupervised learning produced positive outcomes in these learning scenarios. MLPs employ recurrent and feed forward networks. The properties of a multilayer perceptron (MLP) include learning with input-output patterns, universal approximation of continuous nonlinear functions, and advanced network designs with numerous inputs and outputs [10].



Some problems may be observed. Certain neural networks exhibit several local minima, and determining the potential number of neurons required for a given task is another factor that influences the network's optimality. It's also important to keep in mind that, despite the neural network solutions' tendency to converge, a unique solution might not emerge [11].
Let's now examine another scenario in which the data is plotted and classified. We see that there are numerous hyperplanes that can be used for classification. Which is superior, though?

Proposed Method:



The following graphic shows how the data are separated by several linear classifiers, or "hyper planes." But only one of them succeeds in achieving the greatest isolation. We need it because, if we use a hyper plane for classification, it might wind up closer to one set of datasets than others. This is something we do not want to happen, which is why the maximum margin classifier or hyper plane notion appears to be the answer. The greatest margin classifier example in the next figure offers a solution to the aforementioned issue [8].

From above illustration, there are many linear classifiers (hyper planes) that separate the data. However only one of these achieves maximum separation. The reason we need it is because if we use a hyper plane to classify, it might end up closer to one set of datasets compared to others and we do not want this to happen and thus we see that the concept of maximum margin classifier or hyper plane as an apparent solution. The next illustration gives the maximum margin classifier example which provides a solution to the above mentioned problem [8].

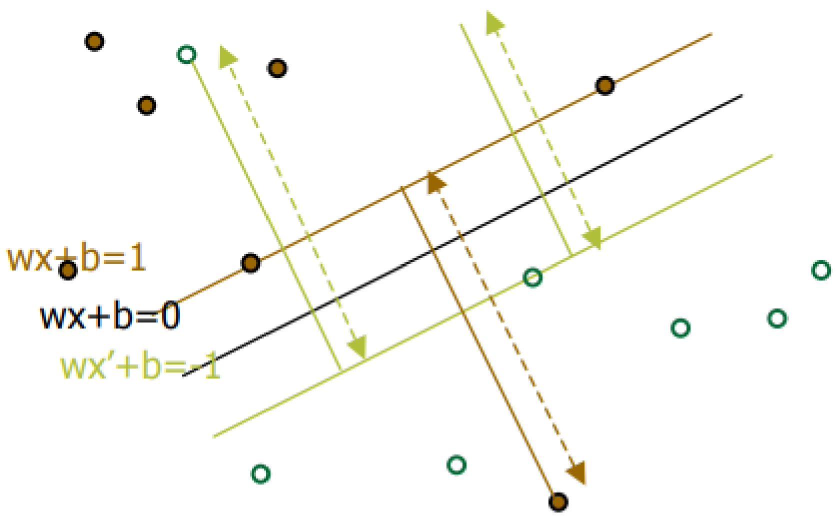


Figure In Hyper Plane

he purpose of the kernel function is to make it possible to carry out operations in the input space as opposed to the perhaps high dimensional feature space. Therefore, there is no need to analyze the inner product in the feature space. The function's task is to transfer the input space's properties to the feature space. A key component of SVM's functionality is the kernel function. Reproducing Kernel Hilbert Spaces is its foundation [8] [14] [15] [18].

Minimum Air Pressure	Maximum Air Pressure	Minimum Vapor Pressure	Maximum Vapor Pressure	Minimum Humidity	Maximum Humidity	Maximum Temperature	Minimum Temperature
998.4	998.7	32.6	34.2	71	76	33.2	26.8
995.5	996.8	34.0	33.6	79	80	34.3	36.4
992.7	997.7	34.4	32.8	75	89	33.5	25.6
997.0	1000.7	32.4	31.1	079	095	31.4	25.4
999.5	1002.7	31.9	30.5	83	93	26.6	25.5
1004.4	1003.0	29.3	31.9	077	095	30.7	25.0
992.7	997.7	34.4	32.8	75	78	26.5	25.5
--------	--------	-------	-------	-------	-------	-------	-------
---------	--------	-------	--------	---------	--------	--------	-------
998.5	1003.0	29.3	31.9	077	095	31.4	25.4
997.5	999.0	33.6	25.5	82	90	26.4	24.2
996.2	996.4	32.6	33.0	96	87	31.0	25.6
993.9	996.7	33.4	33.4	86	86	31	25
996.7	1001.2	32.8	33.8	78	80	32.6	25.2

Calculated Minimum Air Pressure:

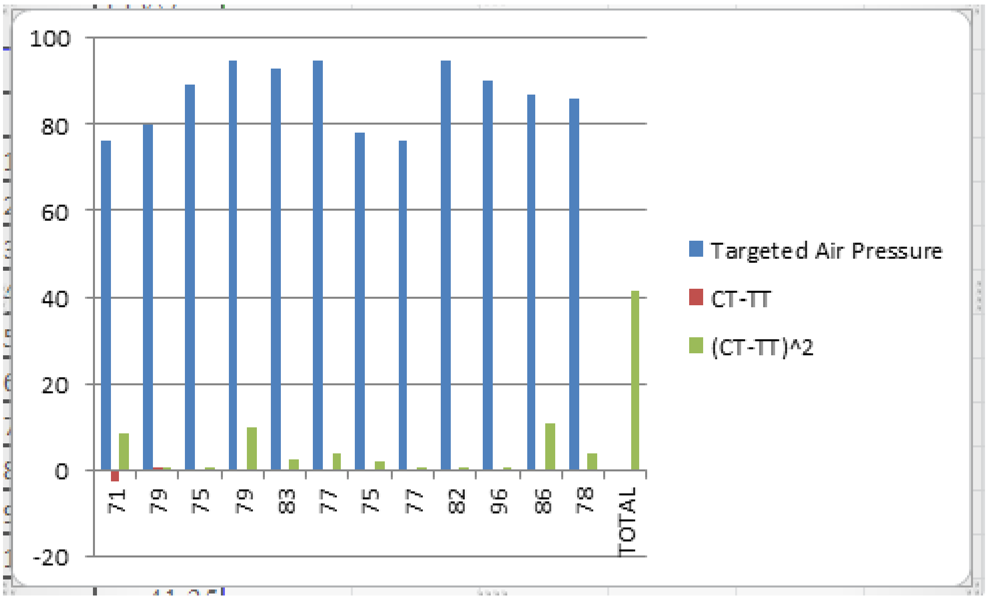
Calculated Humidity	Targeted Humidity	CT-TT	(CT-TT)^2
71	76	-2.9	8.41
79	80	0.7	0.49
75	89	.0.1	0.01
079	095	.0.2	9.61
83	93	.0.3	2.25

077	095	.0.4	3.61
75	78	.0.5	1.69
077	76	.0.6	0.49
82	095	.0.7	0.04
96	90	.0.8	0.25
86	87	.0.9	10.89
78	86	.0.10	3.61
TOTAL			41.35

Now Root Mean Square Formula

$\sqrt{(CT-TT)^2/(N)}$

=1.856



The mathematical Formula $w_1x_1 + w_2x_2 + \cdots + w_nx_n - \theta = 0$

Calculated Air Pressure	Targeted Air Pressure
94.1	095
92	93
96	095
77	78
077	76
93	095
92	90
86	87
85	86
94	095
32	93
93	095

Total=28.23	

Now Root Mean Square Formula

√ (CT-TT)^2/(N)

=1.533

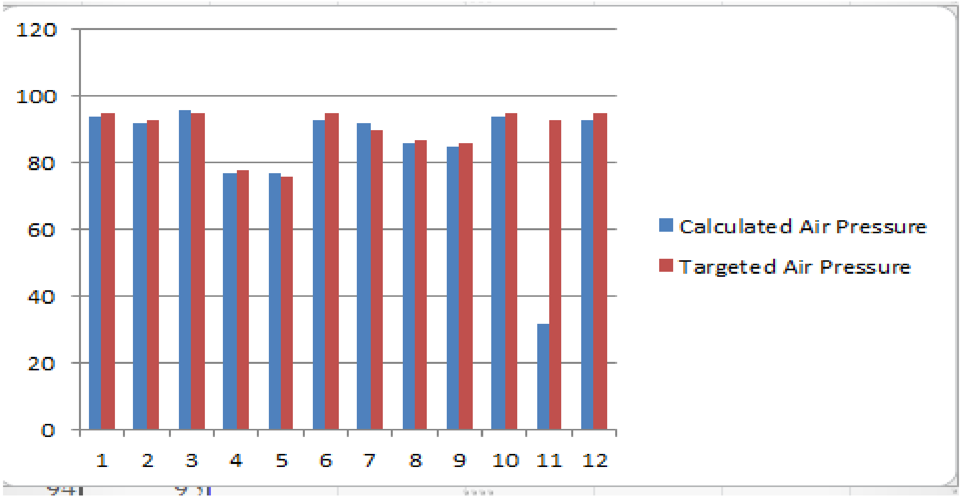


Figure 2: Calculated and Targeted Minimum Air Pressure

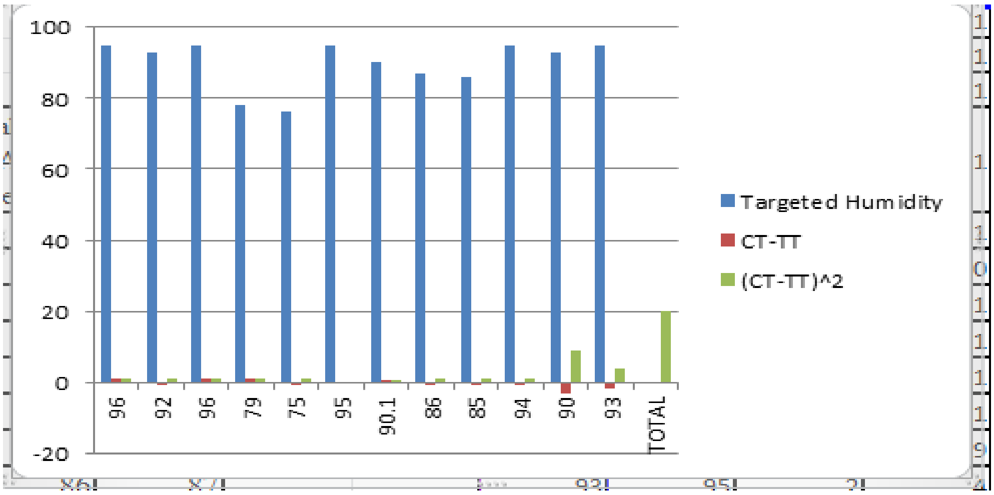
Calculated Maximum Humidity:

Calculated Humidity	Targeted Humidity	CT-TT	(CT-TT)^2
96	95	1	1
92	93	-1	1
96	95	1	1
79	78	1	1
75	76	-1	1
95	95	0	0
90.1	90	0.1	0.01
86	87	-1	1
85	86	-1	1
94	95	-1	1
90	93	-3	9
93	95	-2	4
TOTAL			20.01

Now Root Mean Square Formula =

√ (CT-TT)^2/(N)

=1.29



Calculated Humidity	Targeted Humidity	TT-CT	(TT_CT)^2
75	76	-1	1
79	80	-1	1
88	89	-1	1
95	95	0	0
92	93	-1	1
96	95	1	1
77	78	-1	1
77	76	1	1
75	76	-1	1
91	90	1	1
86	87	-1	1
75	76	-1	1
			11

Now Root Mean Square Formula =

$\sqrt{(CT-TT)^2/(N)}$

=0.957

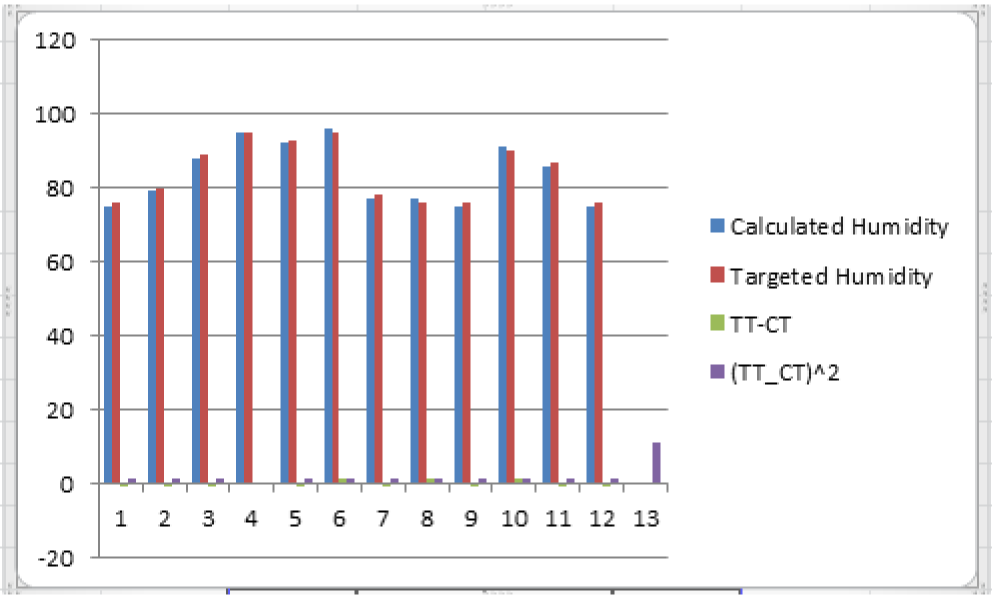


Fig3: Calculated Highest Air Pressure Fig4: Targeted Highest Air Pressure

Support Vector Machine:

Support Vector Machines (SVMs) are versatile machine learning algorithms that are effective in managing high-dimensional data and nonlinear relationships. SVM algorithms are very effective as we try to find the maximum separating hyperplane between the different classes available in the target feature. SVMs can be used for a variety of tasks, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection.

$Y=w^T X+b$

$Y= w^T X+b>=0$

$Y= w^T X+b<=0$

After Applying Support Vector Machine We get the Result

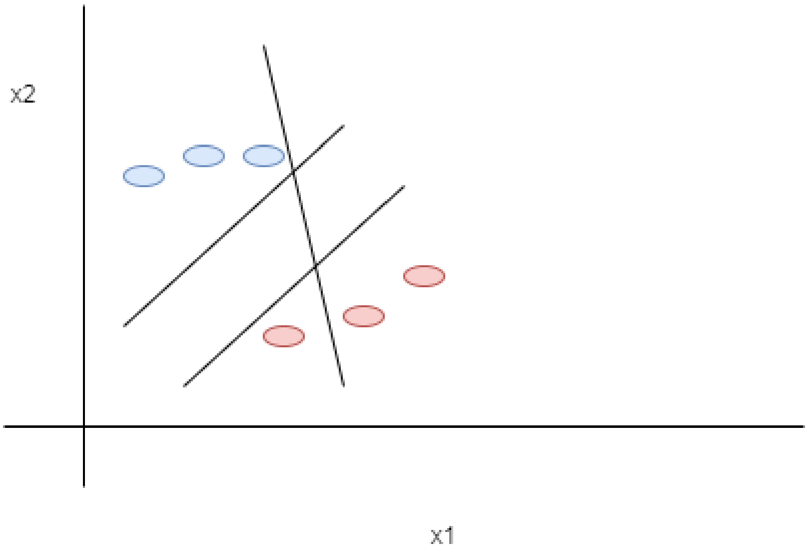


Fig 5: Support Vectors in Hyper Plane

Results:

Minimum Air Pressure:

Calculated Air Pressure	Targeted Air Pressure	CA-TA	(CA_TA)^2
75	76	-1	1
79	80	-1	1
88	89	-1	1
95	95	0	0
92	93	-1	1
96	95	1	1
77	78	-1	1
TOTAL			

Error=0.5354

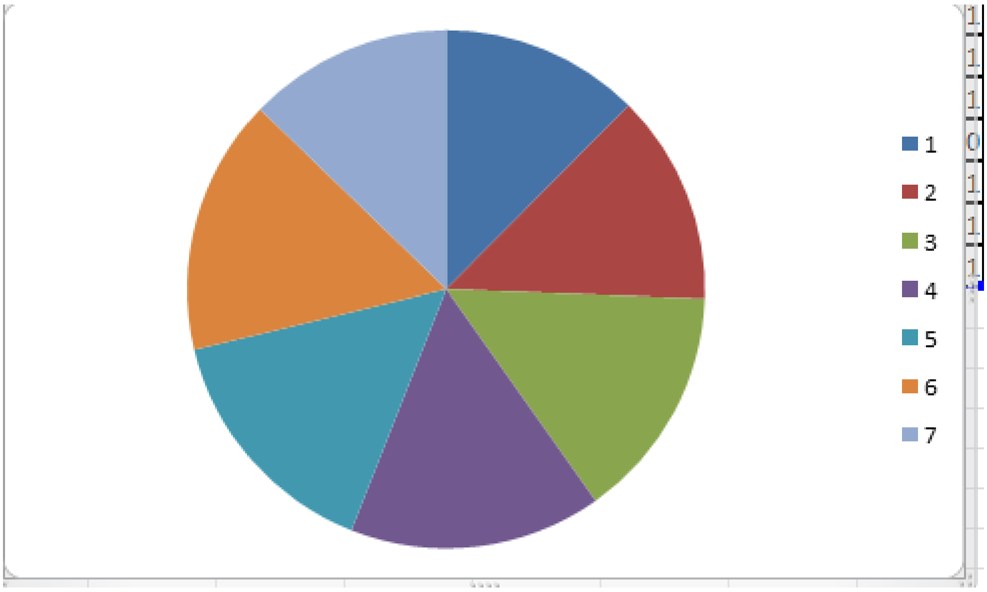
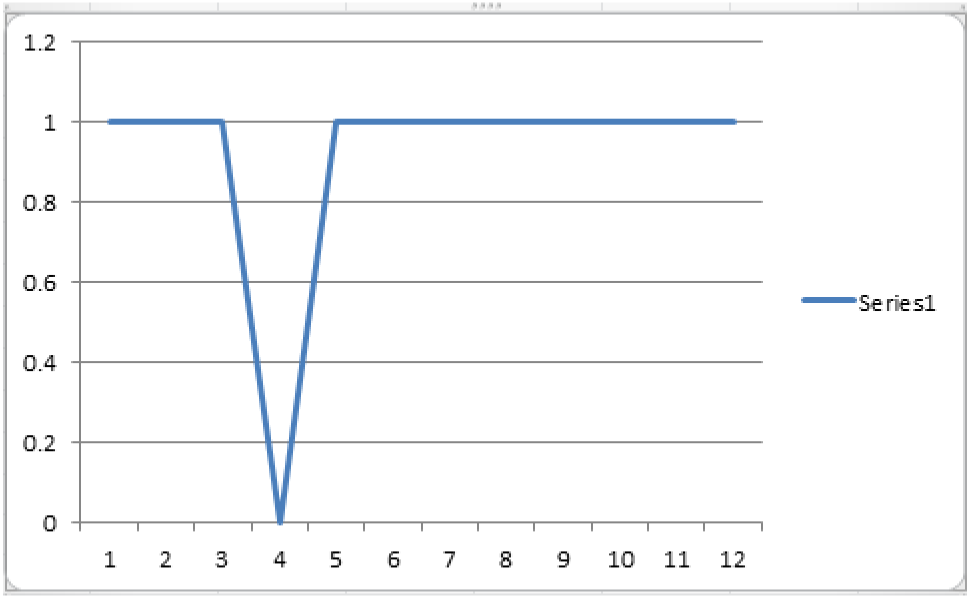


Fig:6 a) Calculted Humidity

Support Vector Machine Application

TT-CT	(TT_CT)^2
-1	1
-1	1
-1	1
0	0
-1	1
1	1
-1	1
1	1
-1	1
1	1
-1	1
-1	1



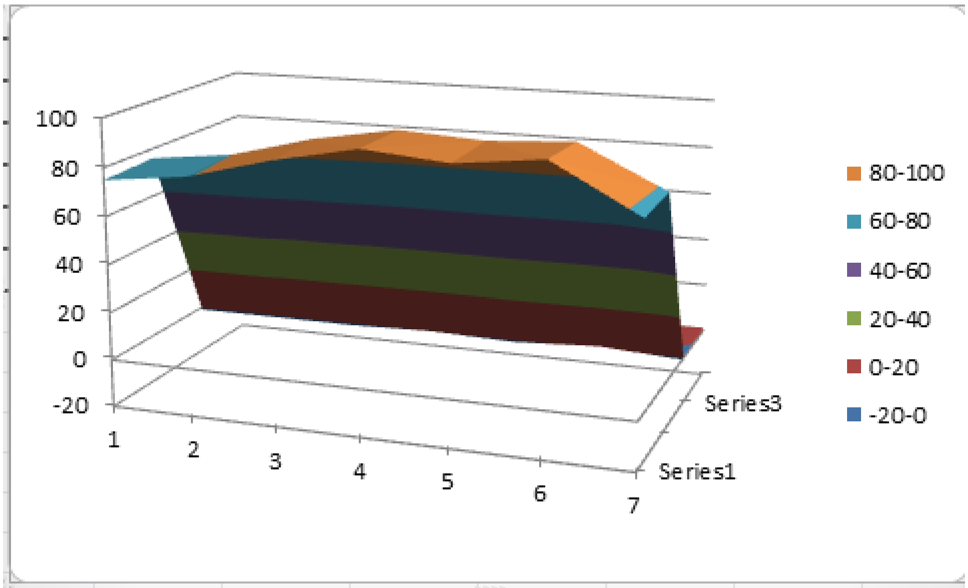


Fig 7: Hyper Plane is giving less Variance for Minimum Humidity

Calculated Air Pressure	Targeted Air Pressure	CA-TA	(CA_TA)^2
96	95	1	1
92	93	-1	1
96	95	1	1
79	78	1	1
75	76	-1	1
95	95	0	0
90.1	90	0.1	0.01
TOTAL			

Error=0.9949

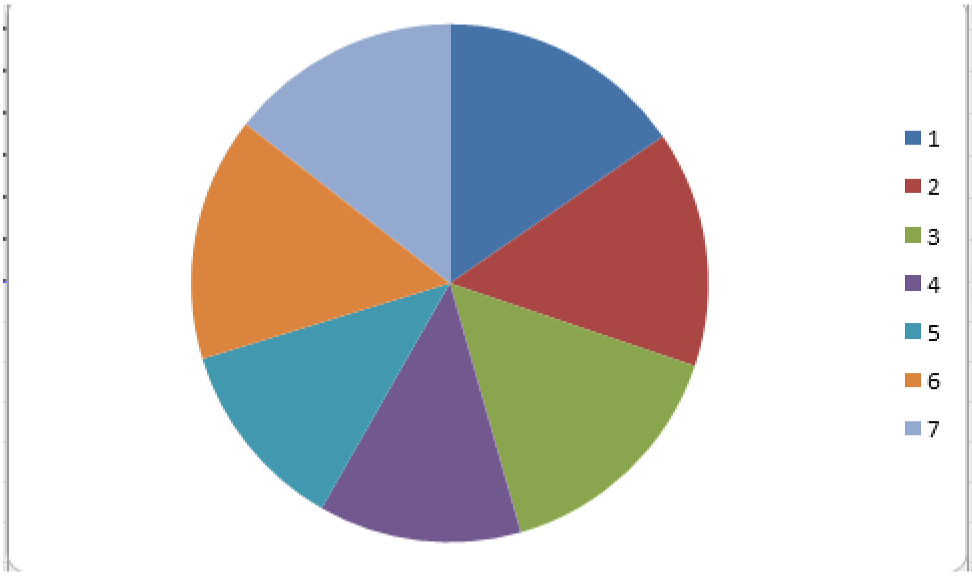
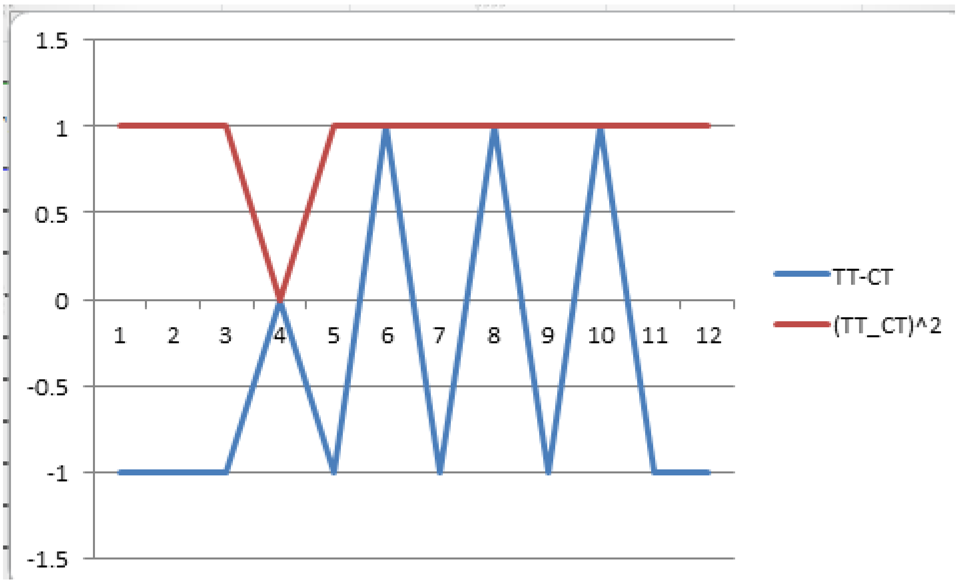


Fig: 8(a) Calculated Maximum and Minimumj Humidity

TT-CT	(TT_CT)^2
-1	1
-1	1
-1	1
0	0
-1	1
1	1
-1	1
1	1
-1	1
1	1
-1	1
-1	1



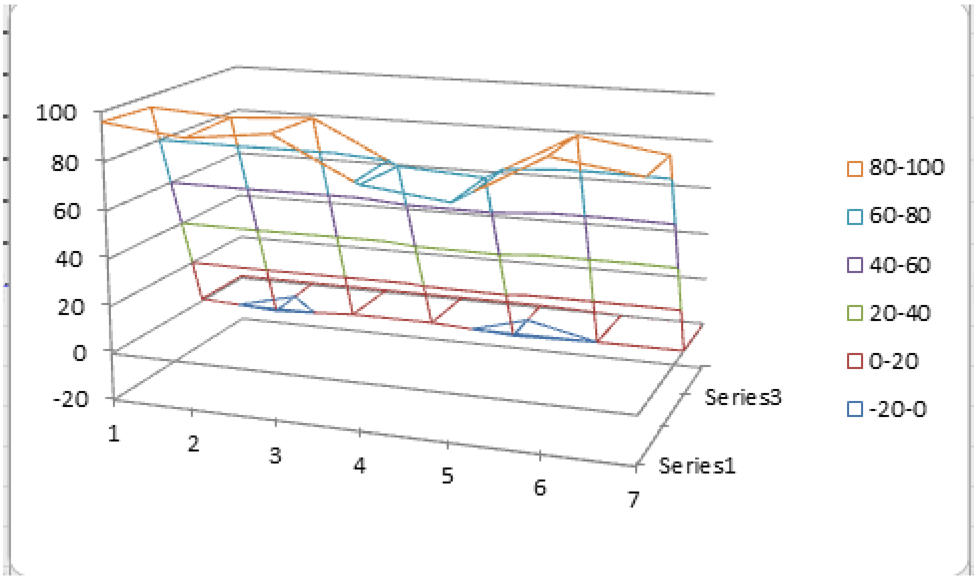


Fig:9: Hyper Plane is giving less Variance for Maximum Humidity

Conclusion:

Support Vector Machine is used instead of Multi Layer Perceptron is used to attain the goal of identifying the lowest and highest Air Pressure together with its approximated model. The importance of the lowest and maximum Air Pressures for the weather is well known. After figuring out the lowest and maximum Air Pressures, we used root mean square to get the minimum prediction error, which will yield results that are roughly accurate.

Reference:

1.Scott H. Brown, Auburn University Montgomery," Multiple Linear Regression Analysis: A Matrix Approachwith MATLAB" ,Alabama Journal of Mathematics, Spring/Fall 2009,PP.1-3

2. T. Soares dos Santos , D. Mendes , and R. Rodrigues Torres," Artificial neural networks and multiple linearregression model using principal components to estimate rainfall over South America", Nonlin. ProcessesGeophys., 23, 2016 PP. 13–20,

3. Katerina Tsakiri ,, Antonios Marsellos and Stelios Kapetanakis," Artificial Neural Network and MultipleLinearRegression for Flood Prediction in Mohawk River,New York", 29 August 2018,PP.1-20

4. Ilaboya, I.R and Igbinedion O. E," Performance of Multiple Linear Regression (MLR) and Artificial NeuralNetwork (ANN) for the Prediction of Monthly Maximum Rainfall in Benin City, Nigeria", INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE AND APPLICATION Ilaboya et al. ,Vol. 3, No. 1, 2019,PP-21-37

5. Brian J. Cerruti and Steven G. Decker ," USING MULTIPLE LINEAR REGRESSION TO DEVELOP A PLANTDAMAGE MODEL FOR A MAJOR UTILITY COMPANY", J11.2,PP.1-6

6. S. Prabakaran, P. Naveen Kumar and P. Sai Mani Tarun," RAINFALL PREDICTION USING MODIFIED LINEARREGRESSION", ARPN Journal of Engineering and Applied Sciences, VOL. 12, NO. 12, JUNE 2017.PP.3715-3718

7.K. Ramesh, 2R. Anitha and 3 P. SelvaGopal," Linear Regression Based Lead Seven Day Maximum andMinimum Air Air Pressure Prediction in Chennai, India", Research Journal of Applied Sciences, Engineering and Technology 7(11): 2014,March 20, 2014, PP.2306-2310,